通用算法开发说明:

朴素贝叶斯：支持多分类算法

算法参数说明：

|  |  |  |
| --- | --- | --- |
| 参数名称 | 释义 | 传参类型 |
| action | 指令：train/test/predictions | 必传参数 |
| labelColName | 标签列名称 | 对于train/test必传参数，其他不传 |
| split | 原数据集分隔方式 | 必传参数 |
| selectCols | 选择参与计算的列的列名 | 必传参数 |
| dataSetPathFile | 原数据集存放地址 | 必传参数 |
| destPath | 结果数据集存放路径 | 对于test/predictions必传参数，其他不传 |
| allCols | 原数据集表头 | 当原数据集不含表头时必传，否则不传 |
| modelPathFile | 模型存放路径 | test/predictions必传train过程后台生成，并返回给用户 |

接口调用说明：

Eg:原始数据集是.csv文件，在进行模型训练之前，将数据拆分成7：3

在调用train/test/predictions三个接口时，selectCols必须完全保持一致

训练接口参数说明：train

原始数据集的70%用作训练

eg:

{

"labelColName": "label ",

"split": "\u0001",

"action": "train",

"selectCols": [ "A", "B","C","D","E"],

"allCols": ["A","B","C","D","E","F", "G"," label "],

"dataSetPathFile": "/xx/xx/xx/train"

}

模型评估接口参数说明:test

原始数据集的30%用作测试

{

"labelColName": "label ",

"split": "\u0001",

"action": "test",

"selectCols": [ "A", "B","C","D","E"],

"allCols": ["A","B","C","D","E","F", "G"," label "],

"modelPathFile": "model15160924315864cac33d5bab4417fb93ea912ba9e8f4c",

"destPath": "/xx/xx/xx/result"

}

eg:

预测接口参数说明:predictions

对新的数据进行预测：

{

"split": "\u0001",

"action": "predictions",

"selectCols": [ "A", "B","C","D","E"],

"allCols": ["A","B","C","D","E","F", "G"," label "],

"modelPathFile": "model15160924315864cac33d5bab4417fb93ea912ba9e8f4c",

"dataSetPathFile": "/xx/xx/xx/prediciton",

"destPath": "/xx/xx/xx/result"

}